Two Multiple Regression Models of Small-Area Population Change

CHARLES H. GRAVES
Planning Consultant, Puget Sound Regional Transportation Study, Seattle

•IN THE last few years increased attention has been given to the development of models which explain the spatial distribution of population within an urban region. Models have been constructed utilizing multiple regression, linear and nonlinear programming, dynamic programming, calculus, and probability theory (1, 2). The impetus for this research activity has come largely from improvements in computer technology, from regional transportation studies requiring small-area population forecasts to carry out travel forecasting procedures, and from community renewal programs concerned with plans for small areas. A challenge in building models of population distribution designed to be applied to travel forecasting comes from the very small size of the forecast areas, which typically range from several square blocks to a few square miles. A very sensitive model is required to allocate population change accurately to such small areas.

DEVELOPMENT AND TESTING OF MODELS

This paper describes the development and testing of two linear multiple regression models used to explain and to predict the distribution of population change in small areas inside the Puget Sound region. One model was developed to explain population growth and another to explain population decline. The specific objective of the construction of the population models was to discover the relative influence of certain factors hypothesized to have influenced the distribution of population change in the Puget Sound region during the decade of 1950 to 1960.

The linear multiple regression model took the form:

$$Y = a + b_1 X_1 + b_2 X_2 + b_3 X_3 \ldots\ldots$$

where

Y = index of population change,
a = constant (intercept),
$b_1, b_2, b_3 \ldots$ = the regression coefficients, and
$X_1, X_2, X_3 \ldots$ = independent variables hypothesized to have influenced the distribution of population change.

Eleven independent variables were hypothesized as determinants of population change: (a) accessibility to employment, (b) logarithm of accessibility to employment, (c) land availability, (d) size of land parcels under single ownership, (e) income level of resident population, (f) occupation index of resident population, (g) combination of income and occupation, (h) age of housing, (i) condition of housing, (j) combination of age and condition of housing, and (k) lot size permitted by zoning.

Area Studied in Model Development

Population change from 1950 to 1960 was used as the basis for developing the regression models. Census tracts were selected as the basic areal unit, since they represent a unit for which considerable data are available from which to measure many

Paper sponsored by Committee on Land Use Evaluation.

of the independent variables, and because they provided the only reliable source of information for the basic dependent variable, population change between 1950 and 1960.

Tracts which showed population growth, located primarily outside central cities, were studied separately from declining tracts located within the central area of Tacoma because it was believed that a separate set of residential location factors operate on declining areas compared to growing areas, and because of computational advantages. Data were not available for tracts which did not have comparable boundaries in 1950 and 1960. Therefore, these tracts were excluded from the study. A set of 74 tracts exhibiting population growth and 17 tracts of population decline were selected for regression analysis. Figure 1 shows their location. The population change in these tracts represented about one-half of the total population change between 1950 and 1960 in the Puget Sound study area.

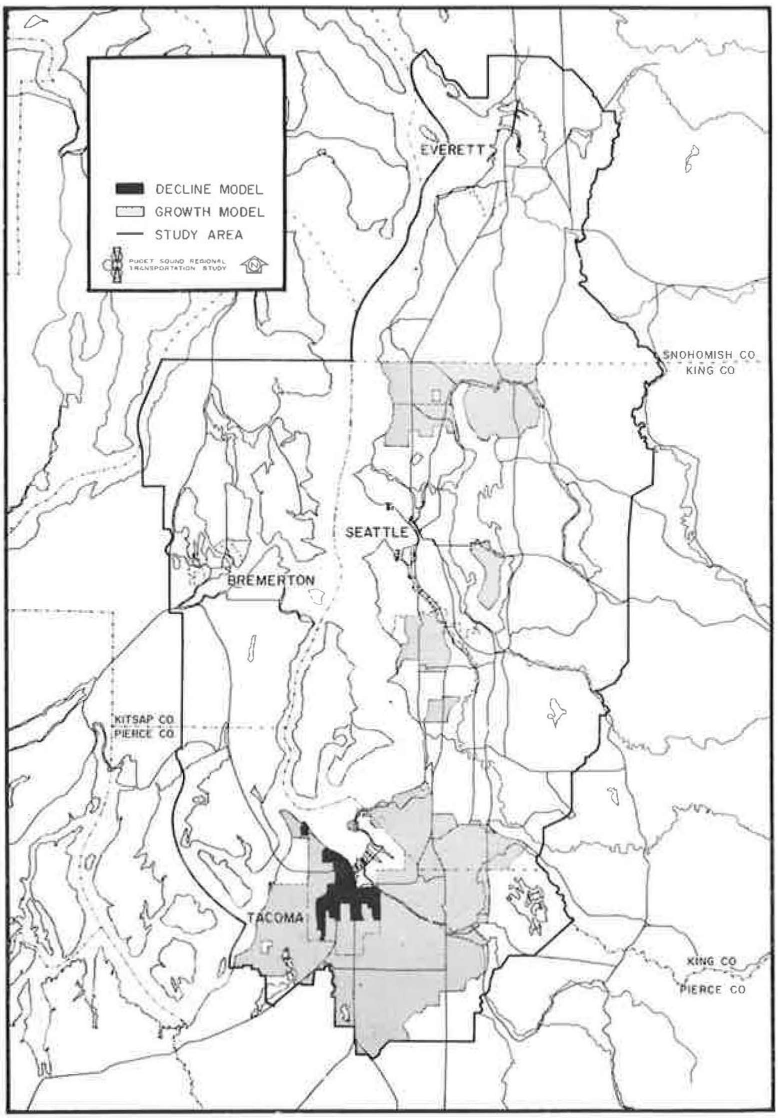

Figure 1. Location of census tract test areas, population model.

Definition of Variables and Their Logic

Six variations of the dependent variable, population change between 1950 and 1960, were chosen for study:

1. Total population change;
2. Ratio of actual to hypothetical total population change;
3. Ratio of actual to hypothetical change in number of single-family dwelling units;
4. Logarithmic transformation of 1;
5. Logarithmic transformation of 2; and
6. Logarithmic transformation of 3.

Variable 1 is a straightforward statement of the dependent variable. Variable 2 is the ratio of the actual to hypothetical population change, where the hypothetical change is that which would have resulted if residential land availability were the only determinant. For example, if a study tract had 10 percent of the additional land available in the area studied, then the hypothetical change was assumed to be 10 percent of the total population change in the area. An attempt was then made to explain the difference between the actual and the hypothetical change by regression against independent variables other than land availability. Dependent variable 3, the ratio of the actual to a hypothetical change in the number of single-family dwelling units, was tested to see if a better model could be built for low-density population than for population as a whole. The hypothetical change was analogous to that developed for variable 2. Dependent variables 4, 5, and 6 are common logarithmic transformations of variables 1, 2, and 3.

The independent variables selected for study were intuitively believed to determine the spatial distribution of population. The independent variables were limited to those which had an intuitive basis to safeguard against the risk of finding specious correlations.

The following eleven independent variables were tested:

1. Accessibility to Employment—Accessibility to employment measures the collective desire on the part of people to locate their residences in relation to employment so as to minimize their work trips, and also their desire to minimize transportation costs to the whole range of urban activities such as shopping centers and office complexes which are also represented by employment. Accessibility to employment was computed over the 1961 highway network to 1961 employment using a gravity-type formula:

$$A = \frac{S_1}{T_{1-1}{}^X} + \frac{S_2}{T_{1-2}{}^X} \cdots \frac{S_n}{T_{1-n}{}^X}$$

where

A_1 = index of accessibility of zone 1 to employment in all other zones,
S_1 = number of employees in zone 1,
T_{1-2} = travel time (including terminal time) between zones 1 and 2, and
X = exponent representing tripmakers' resistance to distance.

A travel time exponent of 2 was used. Employment data used in the equation were derived from 1961 first work trip information collected as a part of the study's origin-and-destination survey. Travel time was computed over the 1961 highway network. The equation produced values representing 1961 accessibility. Independent variables, however, must be considered to be operating at the beginning of the study period, in this case 1950. Travel time and employment data for 1950 were not available and, therefore, it was necessary to assume that 1961 and 1950 values were relatively the same. This probably is a reasonable assumption since the pattern of employment in the Puget Sound area did not change much between 1950 and 1960, and transportation improvements tended to be minor and scattered.

2. Logarithm of Accessibility to Employment—Accessibility to employment was transformed to its common logarithm to replicate Hansen's regression analysis (3).

3. Vacant Residential Land Available—Land availability was defined as the additional holding capacity of land in 1950. Holding capacity was derived from the measurement of the amount of vacant, developable land in 1960. This was defined as vacant land zoned for residences or unzoned, less deductions for the areas of excessive slopes, poor soils, floodplains, and public ownership. Land area was converted to 1960 holding capacity using a density factor derived from the density permitted by zoning and data on average lot sizes. Holding capacity in 1960 was converted to 1950 capacity by adding population change between 1950 and 1960. The logic behind the test of this variable is simply that the chance for residential development increases as the amount of available land increases.

4. Size of Land Parcels in Single Ownership or Control—This variable was defined using land availability as a proxy since it was not feasible to collect actual data about ownership sizes from assessors' or other records.

5, 6, and 7. Income, Occupation, and Their Combinations—An index of income and occupation in 1950 was developed from U. S. Census data. The index of median income of families and related individuals for each of the census tracts being studied was defined as the ratio of the median income of that tract to the median income of the Standard Metropolitan Statistical Area. The occupation index was defined as the ratio of white-collar workers to all workers. White-collar was defined as professional, technical, and kindred workers; managers, officers, and proprietors; salesworkers; and craftsmen, foremen, and kindred workers. It was intended to let income or occupation, or both combined, stand for the relative prestige of the tract. The logic behind the test of prestige as a determinant of population distribution is that many people desire to move upward on the prestige scale through the choice of their residential location. High-income, white-collar families are particularly mobile and their movement may in turn attract other lower income and occupation groups. Thus, prestige areas tend to "trigger" new development. Income and occupation indexes were tested separately and in combination on the theory that both are components of prestige.

8, 9, and 10. Age and Condition of Housing and Their Combination—The age of housing was defined as the ratio of housing built before 1929 to total units in 1950, using 1950 census data. Housing condition was defined as the ratio of unsound to total housing units in 1950. It was expected that these variables would be negatively correlated with population growth because the existence of old or unsound housing should have a depressing effect on the area's attractiveness for growth, particularly in a period of rising incomes.

11. Average Net Lot Size—This variable was defined as the average net lot size for single-family units permitted by zoning. It tests the theory that the market demands smaller lot sizes than generally permitted by zoning.

Equations Tested and Analysis

Simple correlation and linear multiple regression analysis were used to select combinations of variables which statistically best explained population growth and decline. The best combination was defined as the set of independent variables whose simple correlations were in the direction hypothesized when regressed against the dependent variables, and which maximized the proportion of explained variance, R^2, subject to satisfactory confidence levels (1 percent or better). An attempt was made to avoid testing interlinked independent variables in the same equation. Two series of equations were tested, one to explain population growth and another to explain population decline.

1. Population Growth Model—The number of observations in the population growth model was 74. Table 1 gives the simple correlations between all of the variables tested. The direction of correlation hypothesized for the independent variables was

TABLE 1

POPULATION GROWTH MODEL: SIMPLE CORRELATIONS

INDEPENDENT VARIABLES	DEPENDENT VARIABLES: Direction Hypothesized	Population Change	Actual/Hypothetical Total Population Change	Actual/Hypothetical Single Family Dwelling Unit Change	Logarithm of Population Change	Logarithm of Actual/Hypothetical Total Population Change	Logarithm Actual/Hypothetical Single Family Dwelling Unit Change	INDEPENDENT VARIABLES: Accessibility to Employment	Logarithm of Accessibility	Land Availability	Land Ownership	Income	Occupation	Income & Occupation	Age of Housing	Condition of Housing	Age and Condition
Accessibility to Employment	+	*-.02	.41	.16	.11	.50	.45										
Logarithm of Accessibility	+	*-.01	.48	.19	.11	.56	.52	.98									
Land Availability	-	*.28	-.49	-.31	*.23	-.64	-.56	-.42	-.49								
Land Ownership	-	*.28	-.49	-.31	*.23	-.64	-.56	-.42	-.49	1.00							
Income	+	.27	.52	.17	.25	.55	.47	.43	.48	-.24	-.24						
Occupation	+	.32	.34	.02	.41	.36	.26	.28	.32	-.08	-.08	.38					
Income and Occupation	+	.33	.54	.14	.36	.57	.47	.45	.50	-.22	-.22	.93	.69				
Age of Housing	-	-.34	-.18	*.22	-.54	-.25	-.05	-.48	-.45	-.01	-.01	-.32	-.49	-.44			
Condition of Housing	-	-.15	-.45	-.27	-.15	-.48	-.47	-.36	-.42	.33	.33	-.33	-.28	-.37	.06		
Age & Condition	-	-.36	-.30	*.13	-.55	-.36	-.17	-.53	-.55	.06	.06	-.45	-.51	-.55	.97	.29	
Lot Size	-	*.16	-.35	-.35	*.25	-.27	-.39	-.25	-.26	.25	.25	-.22	.15	-.11	-.25	.40	-.11

* Not in direction hypothesized

confirmed for the regression of three of the six dependent variables tested. These three variables were (a) the actual to the hypothetical total population change, (b) its logarithmic transformation, and (c) the logarithmic transformation of the actual to the hypothetical change in the number of single-family dwelling units. Table 1 indicates that the regression of the remaining three dependent variables produced some simple correlations in conflict with the hypothesis, casting doubt on their validity.

TABLE 2

POPULATION GROWTH MODEL: REPRESENTATIVE COMBINATIONS OF VARIABLES ANALYZED BY MULTIPLE REGRESSION

	Equation No.	#1		#2		#3		#4		#5		#6		#7		#8		#9	
	Variables	S	P	S	P	S	P*	S	P	S	P	S	P	S	P	S	P	S	P
Independent Variables	Accessibility	.21	25			**NS	17			.33	03			NS	00				
	Logarithm of Accessibility			.19	00			.01	06			.01	07			NS	01	.12	01
	Land Available													.01	07	.04	07		
	Land Ownership	NS	17	NS	22	NS	08			NS	02							NS	21
	Income					.01	14			.23	01	.01	22	.01	10	.06	06	.01	30
	Occupation					.05	04			.21	00	.09	03	.25	04	.18	07		
	Income and Occupation	.01	10	.01	18			.01	18										
	Age of Housing					NS	00			.04	07	.05	01	NS	06	NS	10		
	Condition of Housing	NS	09	NS	23	NS	02	NS	17	NS	01	NS	04	NS	03	NS	02	NS	10
	Age & Cond. of Housing																		
	Lot Size	NS	02			NS	04	NS	07	NS	10	NS	09	.23	04	.09	10		
Dependent Variables	Population Change 1950-1960													X					
	Actual/Hypoth. Total Pop. Change					X													
	Actual/Hypoth. Sing. Fam. D. U. Change									X									
	Logarithm of Population Change															X			
	Log. Actual/Hypoth. Total Population Change	X		X				X										X	
	Log. Actual/Hypoth. Single Family D. U. Change											X							
	Coefficient of Determination (R^2)	.6296		.6289		.4989		.4742		.2334		.4673		.3407		.4436		.6141	
	Significance Level (F test)	.01		.01		.01		.01		.05		.01		.01		.01		.01	

*S = Significance Level (t test) and P = Proportion of explained variance
**NS = Not Significant X = Indicates Dependent Variable

The correlations between independent variables are not high enough to indicate substantial interdependence, except where one variable is a combination of two, as one would expect. Factor analysis could be used to test for interdependence.

Table 2 sets out a few of the combinations of variables analyzed by multiple regression using a BIMD 06 program (4). Table 2 also shows for these equations the coefficient of determination, R^2, and significance level, and for each independent variable the proportion of variance explained and significance level. The regression of population change and the change in the number of single-family dwelling units, and their logarithmic transformations, did not meet the R^2 and significance level criteria as well as the dependent variable actual to the hypothetical population change and its logarithmic transformation.

Equation 01 was selected as best because it exhibited a high R^2 and satisfactory significance level; it contained a larger number of independent variables than its close competitors.

Of the 11 independent variables tested, six were discarded and five retained. The five variables retained were (a) access to employment, (b) income and occupation combined, (c) housing condition, (d) lot size, and (e) size of land parcels under single control.

Figures 2 and 3 show the distribution of actual population growth in the selected study tracts between 1950 and 1960 vs the growth estimated by the model. The model explains approximately 63 percent of the variance in population growth.

2. Population Decline Model—Seventeen observations were used to construct the population decline model. The percent of decline in population between 1950 and 1960 was the dependent variable. Seven independent variables were tested. These are the following (with the + and - signs indicating the expected direction of simple correlation): (a) access to employment (-), (b) income level of resident population (-), (c) occupation index of resident population (-), (d) age of housing (+), (e) condition of housing (+), (f) combination of income and occupation (-), and (g) combination of age and condition of housing (+). These variables were defined the same as for the population growth model.

Table 3 gives the simple correlations between all of the variables tested. In the population decline model the hypothesized direction of correlation was confirmed except for age of housing. This variable was, therefore, eliminated from further analysis. No substantial interdependence of variables was indicated, although a factor analysis was not performed to test for interdependence.

Table 4 gives the combinations of variables analyzed; their coefficients of determination, R^2; significance level; and for each independent variable, the proportion of variance explained and the significance level. Eq. 4 appears to be best because of its high R^2 and satisfactory significance level.

Figure 3 indicates the distribution of the actual percent of population decline between 1950 and 1960 vs the decline estimated by the model. About 81 percent of the population decline is explained by the model.

Use of Models to Make Population Distribution Forecasts

The population growth and the population decline models were used to distribute the forecasted 1961 to 1985 study area population change to nearly 600 analysis zones in the region. For analysis zones in areas which grew between 1950 and 1960 the following steps were taken:

1. Values for the independent variables were determined for each analysis zone.
2. These values were substituted into the best equation to give, for each of the population growth zones, the logarithm of the ratio of actual to hypothetical growth. (Hypothetical growth was defined as that which would take place if the capacity of available residential land were the only determinant.)
3. These logarithms were then transformed to numbers and multiplied by the capacity of the available residential land. This produced a growth index for each zone.

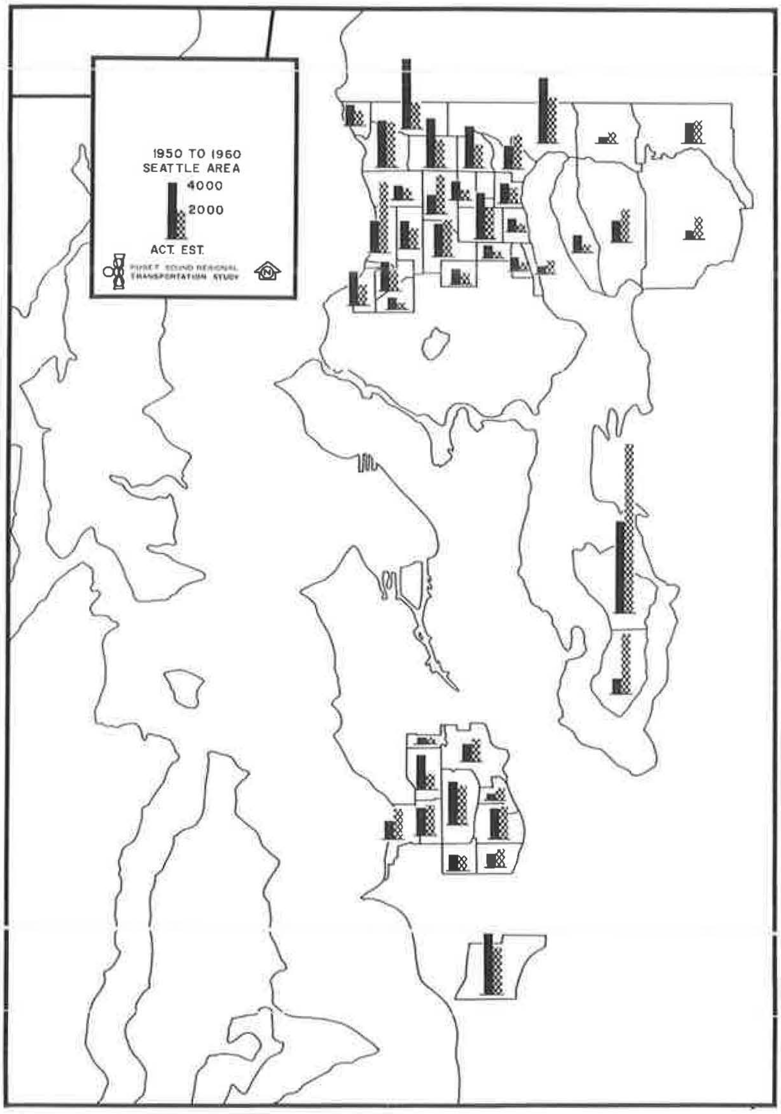

Figure 2. Actual vs estimated change, population models, Seattle area.

4. The 1961 to 1985 population change control totals by county in the study area were distributed to analysis zones by prorating according to the size of the computed growth index.

5. The population distributions to each analysis zone were then checked to assure that they did not exceed the holding capacity of the zones. If the distribution had, in fact, exceeded the holding capacity, the population excess was removed, the zones which were filled to capacity were removed from the group of zones eligible to receive population, and the sum of the excesses was redistributed by repeating Step 4.

The distribution procedure can be stated mathematically as:

$$\frac{P_1}{P_t} = \frac{R_1 AC_1}{R_1 AC_1 + R_2 AC_2 \ldots R_n AC_n}$$

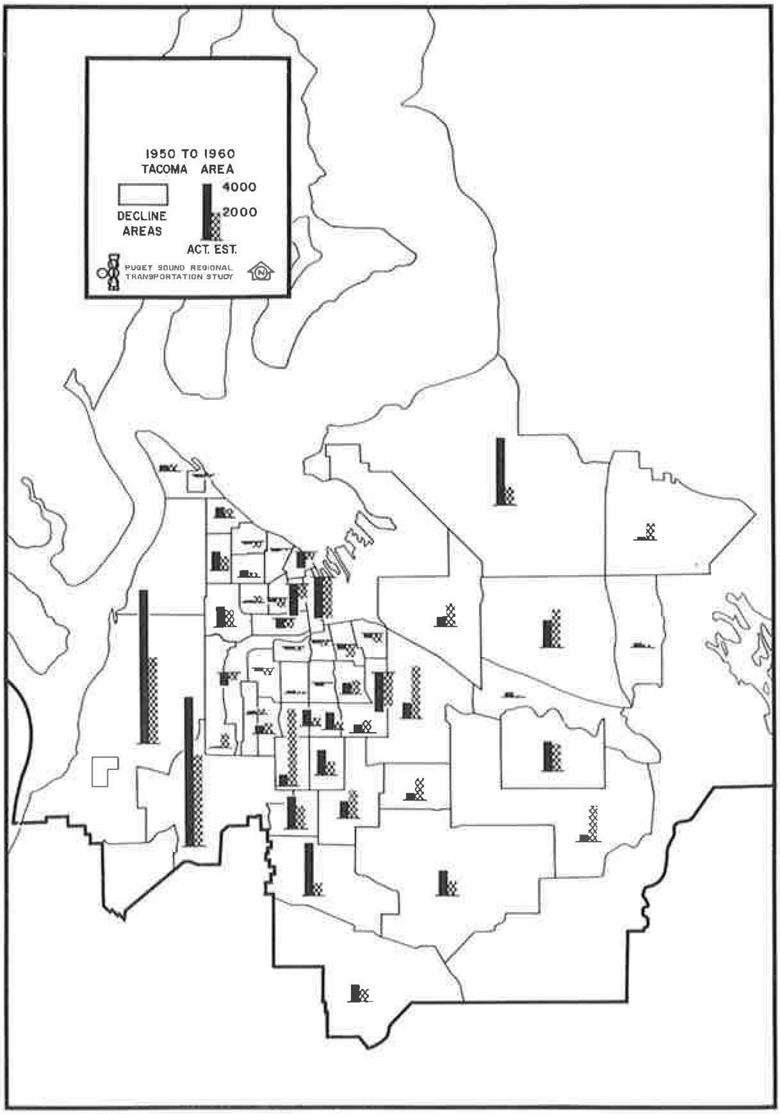

Figure 3. Actual vs estimated change, population models, Tacoma area.

where

P_1 = population change forecast for zone 1 (subject to a holding capacity restraint),
P_t = total population change forecast for the county,
R = ratio of actual to hypothetical change, and
AC = additional holding capacity of developable residential land.

For analysis zones in areas of population decline between 1950 and 1960 the following steps were taken:

1. Values of independent variables were determined for each analysis zone.
2. These values were substituted in the best equation to give for each declining zone the percent decline for a 10-yr period.

TABLE 3

POPULATION DECLINE MODEL: SIMPLE CORRELATIONS

		Direction Hypothesized	% Population Decline 1950-1960	Independent Variables: Access	Income	Lot Size	Occupation	Age of Housing	Condition of Housing	Income-Occupation Combined
	% Population Decrease 1950-1960									
Independent Variables	Access	-	-.05							
	Income		.83	.31						
	Lot Size	+	*-.21	-.38	.30					
	Occupation	-	-.58	-.22	.76	-.02				
	Age of Housing	+	*-.13	.40	-.07	-.65	.22			
	Condition of Housing	+	.63	.02	-.59	-.21	-.46	.34		
	Income-Occupation Combined	-	-.78	-.29	.97	.19	.90	.04	-.57	
	Age-Condition Combined	+	.03	.36	-.21	-.64	.09	.97	.54	-.10

* Not in direction hypothesized

3. The percentages were applied to the 1961 base population to produce a set of estimates of population decline. The total population decline was not set in advance but was aggregated from the results of determining the decline for individual zones.

EVALUATION OF MULTIPLE REGRESSION MODELS AND THEIR APPLICATION TO FORECASTING

This section sets forth some of the advantages and problems associated with the development and use of multiple regression models, and suggests certain avenues of further research. Also, a question is raised concerning the application of the models to forecasting procedures.

TABLE 4

POPULATION DECLINE MODEL: REPRESENTATIVE COMBINATIONS OF VARIABLES ANALYZED BY MULTIPLE REGRESSION

Equation No. / Variables	#1 S	#1 P	#2 S	#2 P	#3 S	#3 P	#4 S	#4 P	#5 S	#5 P
Access							NS	.00	NS	.00
Income	NS	.68	NS	.68			NS	.78		
Occupation			.29	.01			.25	.01		
Condition of Housing	.11	.03	.12	.03	.09	.39	.20	.01	.15	.39
Income-Occupation Combined					NS	.26			NS	.32
% Population Decline	X		X		X		X		X	
R^2	.71		.72		.65		.81		.72	
Significance Level (F)	.01		.01		.01		.01		.01	

S = Significance Level (t Test)
P = Proportion of Explained Variance
NS = Not Significant
X = Dependent Variable

Advantages of Use of Multiple Regression Models

One advantage can be associated with the use of multiple regression to construct population models, as follows:

The multiple regression models in the simple form described in this paper are an operational technique. This quality is important to operating agencies which usually must produce work against severe time and budget limitations. The models described in this paper utilize, for the most part, efficient package programs available from the U.S. Bureau of Public Roads and from the SHARE and BIMD libraries. These package programs keep programming costs low.

The models described in this report share several advantages with some other types of land-use models:

1. Replication or easy updating is possible because all parameters are made explicit.
2. Staff personnel from several agencies can work simultaneously on data collection and model building. Such a joint staff effort encourages effective interagency cooperation.
3. The use of a model and associated computer technology lends an aura of sophistication and correctness to the work, thus facilitating public acceptance of the results.
4. Perhaps the most important advantage of the model is its capacity for projecting population distributions according to novel policy constraints such as those imposed by a regional plan. When an assumption is made of a continuation of past trends and present policies, forecasts produced by handicraft methods based on intuition and experience perhaps may be as accurate as those produced by models. However, when radical or novel assumptions are made about land use, open space and land-use control policies, intuition gained from past experience may not be applicable. When factors, such as the holding capacity permitted by zoning and accessibility to employment over an assumed transportation network, are made explicit parts of a model, these factors can be defined and measured to reflect novel public policies with respect to zoning, employment distributions, and transportation networks. For example, the impact of a proposed bridge or rapid transit system could be translated into accessibility changes which, in turn, could be used to determine the impact of these planned improvements on the population and employment distributions. Similarly, a substantial open space program affects the additional holding capacity of developable residential land, which is one of the independent variables influencing population distributions in the models described in this paper.

Problems with Development of Multiple Regression Models

Certain problems attended the development of the models:

1. One problem is the lack of data or difficulty in developing data. For example, comparable census data were not available for a substantial portion of the study area since it was not tracted by the U.S. Census in 1950. This is likely to be a problem in other urban areas. The accessibility computations require the collection of large amounts of data about employment distributions and operating characteristics of the transportation network which are costly to collect. Regional transportation studies are the only organizations which routinely collect this information.
2. Variables, such as prestige and size of land parcels under single ownership, are difficult to define in a way which accurately reflects the variable to be tested and at the same time permits ease of measurement.
3. Certain problems are presented by the use of multiple regression analysis to develop models. For example, a linear relationship between dependent and independent variables is assumed, there is a danger of interdependence of variables, and the possibility of specious correlations is always present.

Areas for Further Research

Further research may result in improvements in the models described in this paper.

1. Additional variables should be considered for further tests, such as levels of school, sewer, water and gas service; relative tax rates; building code stringency; land value; view amenity; distance from nearest built-up area; and percent built up at the beginning of the study period.

Carroll has tested a promising set of variables with data for the central cities of Minneapolis and St. Paul (5). The dependent variable is percent population change by tract. Independent variables include relative income, percent of dwelling units owner occupied, condition of residential structure, percent of male employed persons in the professional and managerial groups, percent of persons 0 to 19 yr of age, median age of males, average age of structure, relative monthly contract rent, number of persons per room, percent of total dwelling units built in the previous decade, acres of vacant residential land and percent of total dwelling units which are single family. The results of this analysis have not been published. Other variables have been tested by Chapin and Weiss (6). These include marginal land not in urban use, travel distance to the nearest major street, proximity to nonwhite areas, proximity to blighted areas, total travel distance to the high-value corner, proximity to mixed uses, distance to the nearest playground or recreation area, distance to the nearest convenience shopping area, and residential amenity. All of the foregoing variables warrant investigation.

2. The selection of variables would be facilitated by behavioral studies of what motivates the location of homes and employment centers.
3. The systematic mapping of the unexplained variances from regression analysis might uncover additional variables which merit testing (7).
4. Factor analysis should be applied to test for interdependence of variables.
5. Cross-validation should be performed by randomly selecting one-half of the study tracts, constructing a regression equation and predicting the other half, and vice versa.
6. The models should be tested with data from a larger portion of the study area for which a forecast is desired.
7. It might be desirable to build separate models for various parts of the region, such as the central city, suburbia, and exurbia.
8. The equations developed for the Puget Sound region should be tested against data from other regions to determine their universality.
9. Population should be disaggregated by structure type, in a different way than was attempted here, or disaggregated by race, income and other factors.
10. It might be fruitful to experiment with a series of models which explain the migration of population between small zones inside an urban region rather than dealing only with net incremental values.

Problems in the Application of the Model to Forecasting Procedures

Two problems arise in the application of models to long forecast periods in a single step:

1. Values at the beginning of the forecast period should be used for the independent variables but these values can be expected to change over time. The use of a series of short forecast periods rather than one long period permits the values of the independent variables to be updated.
2. The relative weights of the independent variables may be changing over time. The direction and rate of change might be estimated by regressing a set of independent variables against population change for a series of time periods in the past. For example, equations could be constructed to explain population change for 1930 to 1940, 1940 to 1950, and 1950 to 1960. Such an analysis would require historical data which would be costly, difficult, or even impossible to collect.

SUMMARY

This paper has reported the results of research to build two multiple regression models of small-area population change using data from the Puget Sound area. Although the statistical measures of the sensitivity of the models are not very impressive, they produced 1985 forecasts which appeared very good when reviewed by professional planners and engineers familiar with the area's population growth.

A significant advantage in the use of multiple regression as a technique is that it is operational. Another significant advantage of the Puget Sound model, shared by some other models, is its ability to estimate the effect on population distribution of radical or novel land-use and transportation policies.

Finally, there are good prospects for improving the model by carrying out further research along the lines suggested in this paper.

REFERENCES

1. Review of Existing Lane Use Forecasting Techniques. New York, Traffic Res. Corp.
2. Urban Development Models-New Tools for Planning. Spec. Issue, Jour. Amer. Inst. of Planners, May 1965.
3. Hansen, Walter G. How Accessibility Shapes Land Use. Spec. Issue, Jour. Amer. Inst. of Planners, pp. 73-76, May 1959.
4. Bio-Medical Program No. 6, Univ. of California, Los Angeles.
5. Carroll, Don. Land Use Forecasting: Phase 1—Prediction of Dwelling Units For Undeveloped Land. Minnesota Dept. of Highways, unpub. memo, 1963.
6. Chapin, F. Stuart, Jr., and Weiss, Shirley F. Factors Influencing Land Development. Inst. for Res. in Social Sci., Univ. of North Carolina, 1962.
7. Thomas, Edwin N. Maps of Residuals from Regression: Their Characteristics and Uses In Geographic Research. State Univ. of Iowa, Dept. of Geography, No. 2, 1960.